AWARD NUMBER: W81XWH-19-1-0083

TITLE: The Effect of Mecp2 Mutation on Cortical Projections Revealed by Correlated Single-Cell Transcriptomics and Projectomics

PRINCIPAL INVESTIGATOR: Xiaoyin Chen

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

REPORT DATE: January 2021

TYPE OF REPORT: Final report

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
January 2021	Final report	01Apr2019-30Sep2020

4. TITLE AND SUBTITLE

The Effect of Mecp2 Mutation on Cortical Projections Revealed by Correlated Single-Cell Transcriptomics and Projectomics

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-19-1-0083

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Xiaoyin Chen

E-Mail: xichen@cshl.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cold Spring Harbor Laboratory
1 Bungtown Rd,
Cold Spring Harbor, NY 11724

8. PERFORMING ORGANIZATION REPORT NUMBER
61020101

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Rett Syndrome is caused by mutations in *Mecp2*, which result in a constellation of language, cognitive, motor, and autonomic deficits later in life. Although changes in long-range neuronal connectivity likely underlie the behavioral defects in Rett syndrome, it is unclear how long-range axonal projections are disrupted. Here we develop and apply high-throughput single-cell techniques to identify cell type-specific changes in projections in *Mecp2* animals. We identified two subtypes of cortical projection neurons with potential changes in long-range projections, including the corticothalamic neurons and L6b neurons. Our results provide candidate cell types for future in depth studies on the long-range circuitry changes associated with *Mecp2* mutation. Furthermore, our approach is generally applicable to other brain areas and disease models to reveal cell type-specific changes in projections that are difficult to detect using conventional methods.

15. SUBJECT TERMS
Rett syndrome, Mecp2, Long-range projections, high throughput, single cell, MAPseq, BARseq, barcode sequencing, in situ sequencing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	40	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION:

Rett syndrome is a neuropsychiatric disease caused by mutations in the gene *Mecp2*. Although aberrations in long-range neuronal connectivity have been implicated in many neuropsychiatric diseases, such circuitry basis for Rett syndrome is unclear. Such changes in neuronal connectivity likely underlie the behavioral defects in Rett syndrome. In this study, we optimize and use BARseq, a high-throughput technique that maps the projection patterns of neurons by *in situ* barcode sequencing, to identify changes in cortical projections in a cell type-specific manner.

2. KEYWORDS:

Rett syndrome, *Mecp2*, Long-range projections, high throughput, single cell, MAPseq, BARseq, barcode sequencing, in situ sequencing

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Major goal 1: MAPseq analysis of projections in WT and Mecp2- adult mice

Target dates: 7 months

Actual completion dates: 7 months

Major goal 2.1: Technology development of BARseq to correlate projections to gene expression and spatial distribution of neurons

Target dates: 9 months

Actual completion dates: 11 months.

Major goal 2.2: Identify cell type-specific changes in projections caused by Mecp2 mutation

Target dates: 18 months.

Actual completion dates: 18 months.

What was accomplished under these goals?

Major goal 1: Identifying long-range projection defects in Mecp2 mutants

Using MAPseq, we have shown that projection patterns corresponding to corticothalamic neurons are reduced in both the visual cortex and the anterior cingulate cortex (Fig. 1). These results were described in details in the first annual report.

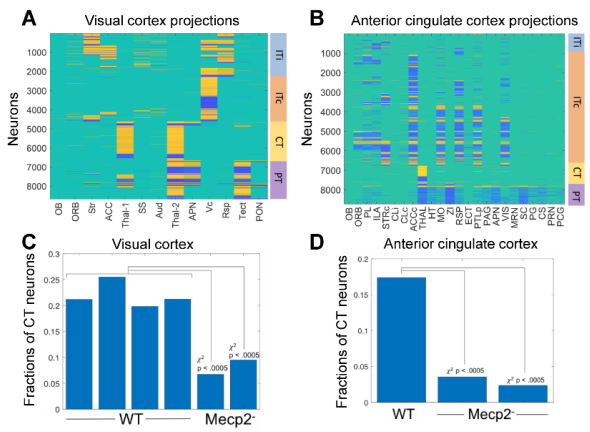

Fig. 1. Projection mapping in Mecp2 -/y and WT animals. (A)(B) Projection matrix of neurons mapped in the visual cortex (A) and anterior cingulate cortex (B). Columns represent projection areas and rows indicate single neurons. Yellow and deep blue bars indicate projections by WT (yellow) and *Mecp2 -/y* (blue) neurons. Neurons are sorted by major classes defined by clustering the projections and color coded by class labels on the right: two classes of intratelencephalic

(IT) neurons with (ITc) or without (ITi) contralateral projections, pyramidal tract (PT) neurons, and corticothalamic (CT) neurons. (C)(D) Fractions of CT neurons in each animal in the visual cortex (C) and the anterior cingulate cortex (D), $p < 0.0005$ comparing each *Mecp2 -/y* brain to each WT brain using Chi-square test.

Major goal 2.1: Technology development of BARseq to correlate projections to gene expression and spatial distribution of neurons

To obtain cell type-specific projection patterns using a barcoding-based approach, we first showed that gene expression in barcoded neurons reflect gene expression in non-barcoded neurons. We then optimized BARseq, a barcoding-based projection mapping technique using *in situ* sequencing, to correlate gene expression, laminar positions, and projections in single cells with high throughput (Fig. 2). These results were described in detail in the first annual report and in Sun et al., 2020, which is currently under revision at Nature Neuroscience.

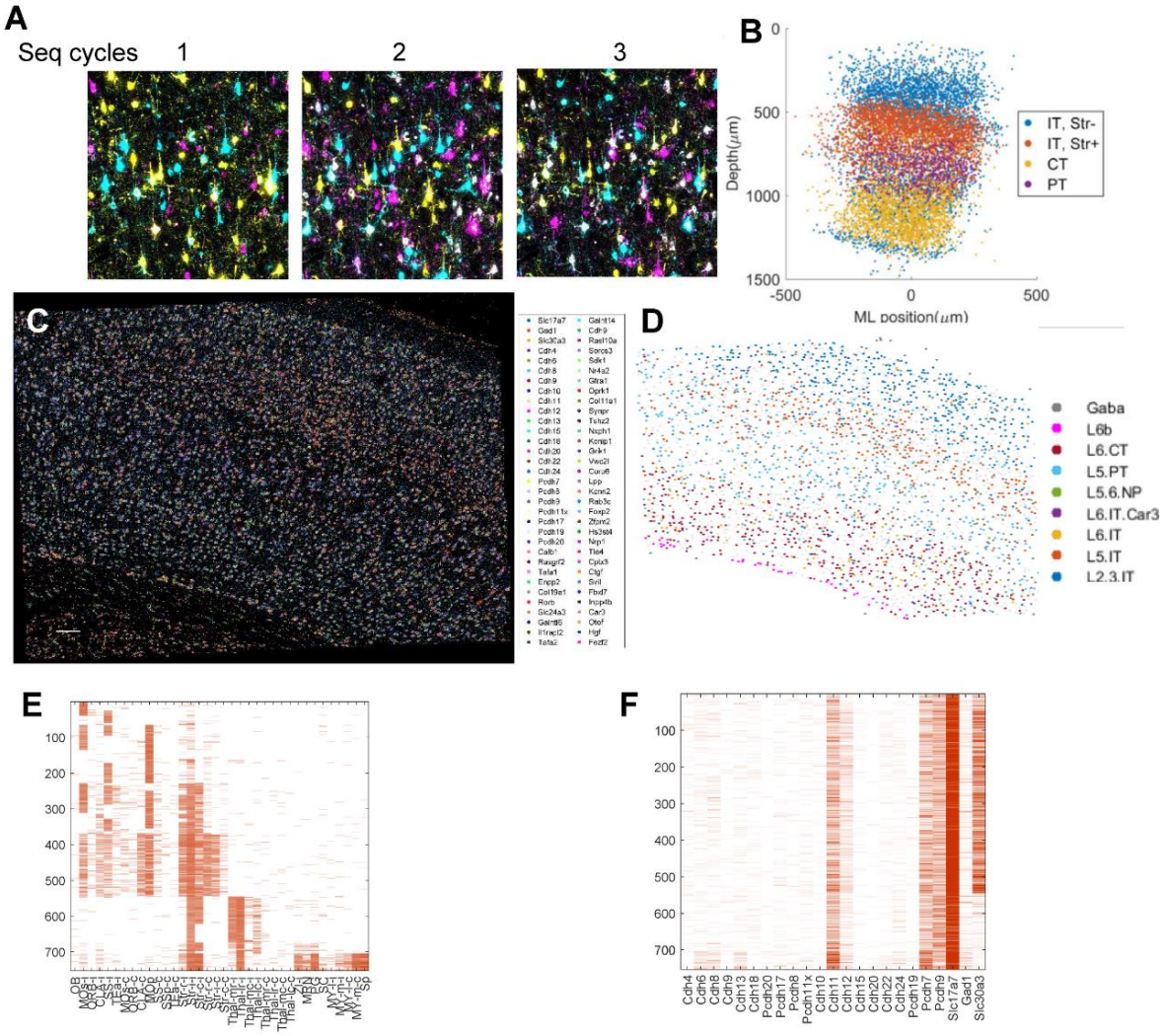

Fig. 2. BARseq correlates projections with gene expression and spatial organization of neurons with high throughput. (A) Three cycles of in situ sequencing images of barcoded neurons in a cortical brain slice. Colors (cyan, yellow, magenta, and white) correspond to the four bases (G, T, A, and C) being read out. (B) the locations of ~10,000 projection neurons in the motor cortex mapped by BARseq. Neurons are color-coded by major clusters of projection patterns. (C) In situ sequencing of 46 endogenous genes that are differentially expressed across excitatory neuronal types in the motor cortex. (D) Transcriptionally defined neuronal types of the same neurons in (C) determined by gene expression. (E)(F) Long-range projections (E) and gene expression (F) of the same neurons in the motor cortex determined by BARseq. Columns indicate projection areas or genes, and rows indicate neurons. Corresponding rows in (E) and (F) represent the same neuron. Neurons are sorted by projection patterns.

Major goal 2.2: Identify cell type-specific changes in projections caused by Mecp2 mutation

We applied BARseq to map the projections of 531 and 181 neurons (Fig. 3a) in the visual cortex from two WT and two *Mecp2 -/y* male animals, respectively. In contrast to the WT animals, the mutant animals lack a population of neurons with only ipsilateral cortical projections at the bottom of layer 6 (Fig. 3b). This population of neurons likely correspond to L6b neurons. This lack of projections could be caused either by defective generation of the neuronal population, or by defective projections of the neurons. To distinguish these possibilities, we performed fluorescent *in situ* hybridization (FISH) in the visual cortex against *Foxp2* and *Cpxl3*, two marker genes specifically expressed in corticothalamic neurons and L6b neurons, respectively. The distributions of these two cell types in the visual cortex of *Mecp2 -/y* animals are comparable to those in WT littermates (Fig. 3c). These results are consistent with the hypothesis that the differences in projection patterns likely reflected changes in projections rather than defective neurogenesis or neuronal survival.

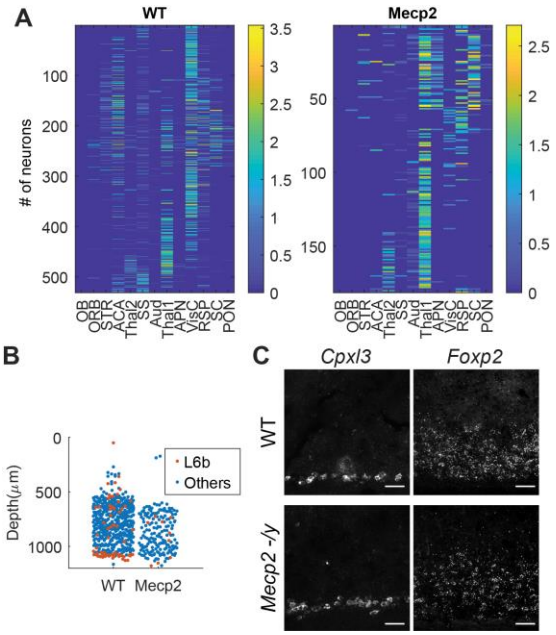

Fig.3 BARseq identifies projection defects in L6b neurons. (A) Projection patterns of neurons in the visual cortex in WT (*left*) and *Mecp2* mutant (*right*) animals. Neurons are sorted by laminar positions. (B) Beeswarm plot of laminar positions of neurons with (red) or without (blue) L6b projection pattern. L6b projection pattern is defined as having projections to ACA or SS, but no projection to VisC and Str. (C) Representative images of FISH against *Cpxl3* (*left*) and *Foxp2* (*right*) in the layer 6 of visual cortex in WT (top) and *Mecp2 -/y* (bottom) animals.

What opportunities for training and professional development has the project provided?

This study represents a significant part of the PI's postdoctoral work. The PI has gotten several job offers for independent research positions, and these works have no doubt contributed to these career advancements.

How were the results disseminated to communities of interest?

The development of BARseq was reported in a preprint (Sun et al., 2020, currently under review), and was presented at the 2021 BRAIN Initiative Investigators Meeting.

What do you plan to do during the next reporting period to accomplish the goals?

Nothing to report (Final report)

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

The finding that projections of corticothalamic neurons and L6b neurons are reduced in *Mecp2 -/y* animals provides a possible neuroanatomical substrate for behavioral defects seen in Rett syndrome. This finding can guide future research in the circuitry mechanisms that lead to these behavioral defects.

The development of BARseq, which allows correlation of gene expression, long-range projections, and spatial organization of neurons, can be further applied to other brain areas to identify changes in gene expression and/or projections associated with *Mecp2* mutation.

What was the impact on other disciplines?

BARseq is generally applicable to other brain areas and disease models. The high throughput and relative low cost of BARseq makes it an attractive tool to screen for connectivity defects in disease models, and to interrogate the organization of nervous systems in general.

What was the impact on technology transfer?

Nothing to report

What was the impact on society beyond science and technology?

Nothing to report

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

We switched to BARseq for cell type-specific projection mapping instead of using single-cell RNAseq and MAPseq. BARseq has higher throughput and can additionally identify the laminar positions of neurons, which is also indicative of cell types. The cost of BARseq per mapped neuron is similar to the original approach of combining single-cell RNAseq and MAPseq. We thus think BARseq is more suitable for achieving the goals of this project.

Actual or anticipated problems or delays and actions or plans to resolve them

Work during the second half of the project was delayed by COVID-19 related shutdowns. More specifically, we were unable to interrogate the developmental trajectory of projection defects in the Mecp2 mutants. We have revised the SOW to reflect these impacts during the last period.

We have switched to using BARseq to correlate projections with gene expression. This switch delayed the completion of Goal 2.1 due to extra work in optimizing BARseq, but the higher throughput of BARseq compared to the originally proposed technique allowed faster completion of Goal 2.2.

Changes that had a significant impact on expenditures

We have switched from using single-cell RNAseq and MAPseq to using BARseq for mapping projections of specific transcriptionally defined cell types. However, this change in approach should not significantly impact the overall cost.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to report

Significant changes in use or care of human subjects

Nothing to report

Significant changes in use or care of vertebrate animals

Nothing to report

Significant changes in use of biohazards and/or select agents

Nothing to report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

  Journal publications.

  Sun, Y.-C., Chen, X., Fischer, S., Lu, S., Gillis, J., Zador, A.M. 2020. Integrating barcoded neuroanatomy with spatial transcriptional profiling reveals cadherin correlates of projections shared across the cortex. bioRxiv, doi: https://doi.org/10.1101/2020.08.25.266460. (*under revision at Nature Neuroscience*)

  Books or other non-periodical, one-time publications. Nothing to report

  Other publications, conference papers and presentations.

  Virtua poster presentation: Shared molecular logic underlying long-range projections across cortical areas revealed by *in situ* sequencing. 2020 BRAIN Initiative Investigators Meeting

- **Website(s) or other Internet site(s)** Nothing to report

- **Technologies or techniques**

  We have improved BARseq, which can correlate gene expression, long-range projections, and spatial organization of neurons with high throughput and cellular resolution. These improvements are reported in Sun et al., 2020. The protocol for BARseq is publicly available at protocol.io: dx.doi.org/10.17504/protocols.io.bdedi3a6

- **Inventions, patent applications, and/or licenses** Nothing to report

- **Other Products**

  We have produced projection data for thousands of neurons at cellular resolution in both wild-type and *Mecp2 -/y* animals. These projection data allow further analysis to understand the organization of cortical projections, which could provide additional insights in connectivity changes in neurological diseases.

  BARseq can be applied to other disease models to understand changes in gene expression and/or long-range projections associated with neurological diseases.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Xiaoyin Chen
Project Role:	PI
Researcher Identifier (e.g. ORCID ID):	0000-0002-2807-6125
Nearest person month worked:	18
Contribution to Project:	Dr. Chen performed and analyzed experiments.
Funding Support:	This award

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to report

What other organizations were involved as partners?

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS

The Quad Chart is submitted in the attachment.

9. APPENDICES:

Sun et al., 2020 is attached.

Integrating barcoded neuroanatomy with spatial transcriptional profiling reveals cadherin correlates of projections shared across the cortex

Yu-Chi Sun[1,*], Xiaoyin Chen[1,*,†], Stephan Fischer[1], Shaina Lu[1], Jesse Gillis[1], and Anthony M. Zador[1,†]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[*]These authors contributed equally

[†]Correspondence: xichen@cshl.edu, zador@cshl.edu

Abstract

Functional circuits consist of neurons with diverse axonal projections and gene expression. Understanding the molecular signature of projections requires high-throughput interrogation of both gene expression and projections to multiple targets in the same cells at cellular resolution, which is difficult to achieve using current technology. Here, we introduce BARseq2, a technique that simultaneously maps projections and detects multiplexed gene expression by *in situ* sequencing. We determined the expression of cadherins and cell-type markers in 29,933 cells, and the projections of 3,164 cells in both the mouse motor cortex and auditory cortex. Associating gene expression and projections in 1,349 neurons revealed shared cadherin signatures of homologous projections across the two cortical areas. These cadherins were enriched across multiple branches of the transcriptomic taxonomy. By correlating multi-gene expression and projections to many targets in single neurons with high throughput, BARseq2 provides a path to uncovering the molecular logic underlying neuronal circuits.

Introduction

Neural circuits are comprised of neurons diverse in many properties, such as morphology (Lin et al., 2018; Winnubst et al., 2019), gene expression (Hodge et al., 2019; Saunders et al., 2018; Tasic et al., 2016; Tasic et al., 2018; Zeisel et al., 2018; Zeisel et al., 2015), and projections (Chen et al., 2019; Han et al., 2018; Harris et al., 2019; Wang et al., 2019). Although recent technological advances have made it possible to characterize the diversity in individual neuronal properties, associating multiple properties in single neurons with high throughput remains difficult to achieve. Investigating the relationship between multiple neuronal properties is essential for understanding the complex organization of neural circuits.

Of particular interest is the relationship between endogenous gene expression and long-range projections in the cortex. Cortical neurons have diverse patterns of long-range projections (Chen et al., 2019; Han et al., 2018; Huang et al., 2020; Oh et al., 2014; Wang et al., 2019) and diverse patterns of gene expression (Hodge et al., 2019; Scala et al., 2020; Tasic et al., 2016; Tasic et al., 2018; Yao et al., 2020; Zeisel et al., 2018; Zeisel et al., 2015; Zhang et al., 2020). The diversity in gene expression can be described by clustering neurons into transcriptomic types, but transcriptomic types have limited power in explaining the diversity of cortical projections beyond the major classes of projection neurons [(Chen et al., 2019; Klingler et al., 2018; Tasic et al., 2018; Zhang et al., 2020), but also see (Economo et al., 2018; Kim et al., 2019)]. The lack of clear correspondence between transcriptomic types and projections in the cortex raises the need to explore and identify other gene correlates of projections, potentially independent of transcriptomic types.

One class of candidate genes that might explain the diversity of projections is the cadherin superfamily, which we will refer to generally as “cadherins.” Cadherins are differentially expressed across cortical layers (Hayano et al., 2014; Krishna et al., 2009; Krishna et al., 2011; Redies, 1997) and cardinal inhibitory cell types defined by transcriptomic and phenotypic characteristics (Paul et al., 2017). Functionally, cadherins are known to specify and maintain neuronal connectivity (Basu et al., 2015; Duan

et al., 2014; Duan et al., 2018; Hayano et al., 2014; Paul et al., 2017; Redies, 1997). Based on these studies, we hypothesize that the expression of specific cadherin superfamily members is correlated with specific patterns of projections (Fig. 1A). To test this hypothesis, we require a high-throughput technique that allows simultaneous multiplexed gene detection with projection mapping to multiple target areas at single-neuron resolution. Although advances in spatial transcriptomics have allowed high throughput and multiplexing capacity, achieving both multiplexed gene detection and high-throughput projection mapping in the same neurons remains difficult.

To achieve high-throughput projection mapping, we recently introduced BARseq (Barcoded Anatomy Resolved by sequencing), a projection mapping technique based on *in situ* sequencing of RNA barcodes (Chen et al., 2019). In BARseq, each neuron is labeled with a unique virally encoded RNA barcode that is replicated in the somas and transported to the axon terminals. The barcodes at the axon terminals located at various target areas are sequenced and matched to somatic barcodes, which are sequenced *in situ*, in order to determine the projection patterns of each labeled neuron. This sequencing-based projection mapping strategy, shared by both BARseq and a related technique, MAPseq (Kebschull et al., 2016), has been repeatedly validated using conventional neuroanatomical methods, and the contribution of potential artifacts—arising from sensitivity, double labeling of neurons, degenerate barcodes, fibers of passage, etc.—has been assessed and quantified in a variety of different systems using multiple techniques (Chen et al., 2019; Han et al., 2018; Huang et al., 2020; Kebschull et al., 2016).

Because BARseq preserves the location of somata with high spatial resolution, in principle it provides a platform to combine projection mapping with other neuronal properties also interrogated *in situ*, including gene expression. We have previously shown (Chen et al., 2019) that BARseq can be combined with fluorescent *in situ* hybridization (FISH) and *Cre*-labeling to uncover projections across neuronal subtypes defined by gene expression. However, these approaches can only interrogate one or a few genes at a time, which would be insufficient for associating a superfamily of cell adhesion molecules to diverse cortical projections.

To identify cadherin superfamily members that correlate with projections in the cortex, we aim to develop a technique to simultaneously map projections to multiple brain areas and detect the expression of dozens of genes in hundreds to thousands of neurons from a cortical area with high throughput, high spatial resolution, and cellular resolution. To achieve this goal, we combine the high throughput and multiplexed projection mapping capability of BARseq with state-of-the-art spatial transcriptomic techniques with high imaging throughput and multiplexing capacity (Ke et al., 2013; Qian et al., 2020). This second-generation BARseq (BARseq2) greatly improves the ability to correlate the expression of many genes to projections to many targets in the same neurons.

As a proof-of-principle, we first demonstrate multiplexed gene detection using BARseq2 by mapping the spatial pattern of up to 65 cadherins and cell-type markers in 29,933 cells. We then correlate the expression of 20 cadherins to projections to up to 35 target areas in 1,349 neurons in mouse motor and auditory cortex. Our study reveals novel sets of cadherins that correlated with homologous projections in both cortical areas. BARseq2 thus bridges transcriptomic signatures obtained through spatial transcriptional profiling with sequencing-based projection mapping to illuminate the molecular logic of long-range projections.

Results

To investigate how cadherin expression relates to diverse projections, we developed BARseq2 to combine high-throughput projection mapping with multiplexed detection of gene expression using *in situ* sequencing (Fig. 1B, C). BARseq2 is based on BARseq, which achieves high-throughput projection mapping by *in situ* sequencing of RNA barcodes (Chen et al., 2019). In BARseq (Fig. 1C, *left*), RNA barcodes are reverse-transcribed and hybridized with a padlock probe that is complementary to the region flanking the barcode region. The barcode sequence is then copied into the padlock probe by a DNA polymerase, effectively "gap-filling" the padlock, and is subsequently ligated. After rolling circle amplification of the circularized padlock, the amplified RNA barcodes are then sequenced *in situ* using Illumina sequencing chemistry and matched to barcodes at target areas to identify projections (Fig. 1B). Projection patterns observed using RNA barcoding are consistent with those obtained using conventional neuroanatomical techniques in multiple circuits, including the locus coeruleus (Kebschull et al., 2016), auditory cortex (Chen et al., 2019), visual cortex (Han et al., 2018), and interregional connectivity across the whole cortex (Huang et al., 2020). Possible technical concerns, including distinguishing fibers of passage from axonal termini, sensitivity, double labeling of neurons, and degenerate barcodes, have previously been addressed and will not be discussed in detail again here. In particular, BARseq in both auditory cortex (Chen et al., 2019) and motor cortex (Chen et al., unpublished observations) produced single-cell projection patterns consistent with conventional retrograde tracing and single-cell tracing, while achieving throughput at least two to three orders of magnitude higher than the current state-of-the-art single-cell tracing techniques. For example, we were able to map up to 5,000 neurons per person-week (Chen et al., unpublished observations), which allowed us to uncover organizational principles of projections that would have been difficult to discover using smaller datasets. Combining barcoded single-cell projection mapping with *in situ* detection of endogenous mRNAs exploits this unique advantage in throughput to efficiently interrogate both neuronal gene expression and long-range projections simultaneously.

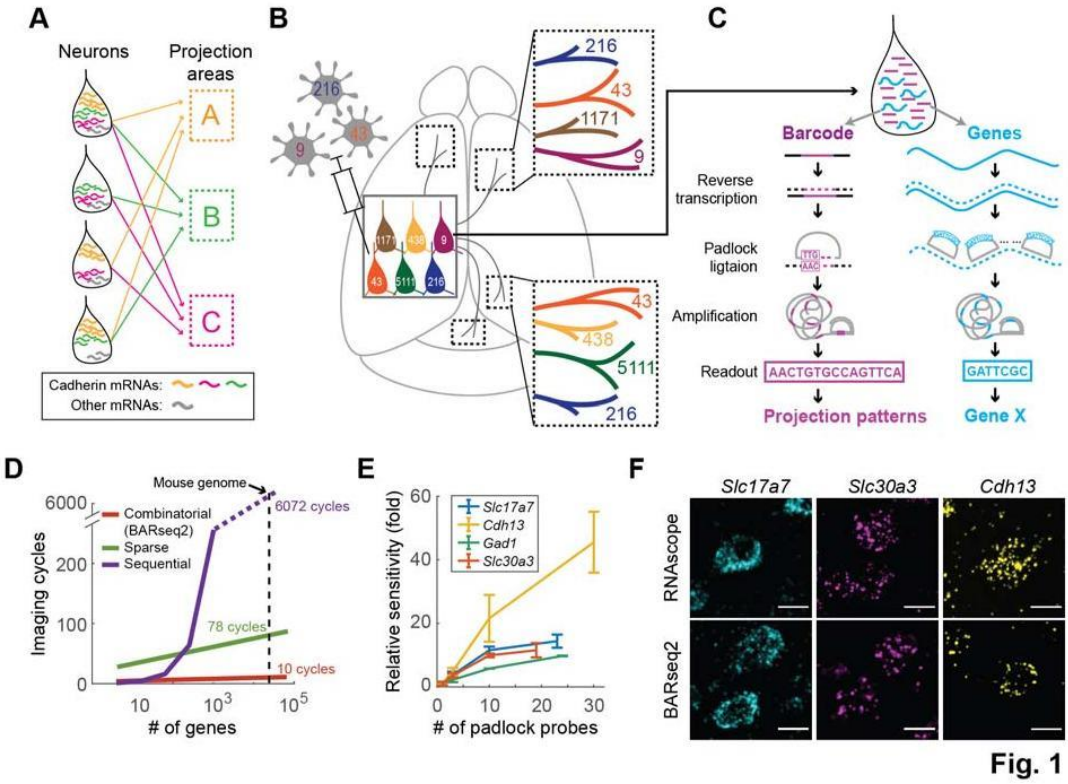

Fig. 1

Figure 1. *In situ* sequencing of endogenous mRNAs using BARseq2. (A) Working hypothesis: the expression of cadherin mRNAs correlates with projections to diverse brain areas. (B)(C) BARseq2 correlates projections and gene expression at cellular resolution. In BARseq2, neurons are barcoded with random RNA sequences to allow projection mapping, and genes are also sequenced in the same barcoded neurons. RNA barcodes and genes are amplified and read out using different strategies (C). (D) Theoretical imaging cycles using combinatorial coding (BARseq2), 4-channel sequential coding, or 4-channel sparse coding as used by Eng et al. (2019). Imaging cycles assumed 3 additional cycles for BARseq2, 1 additional round for sparse coding, and no extra cycle for sequential coding for error correction. (E) Mean ± standard error of the relative sensitivity of BARseq2 in detecting the indicated genes using different numbers of padlock probes per gene. The sensitivity is normalized to that using one probe per gene. n = 2 slices for each gene. (F) Representative images of BARseq2 (*bottom*) detection of the indicated genes using the maximum number of probes shown in (E) compared to RNAscope (*top*). Scale bars = 10 μm.

To detect gene expression using BARseq2, we used a non-gap-filled padlock probe-based approach to amplify target endogenous mRNAs (Ke et al., 2013; Qian et al., 2020)(Fig. 1C, *right*). In this approach, the identity of the target is read out by sequencing a gene-identification index (GII) using Illumina sequencing chemistry *in situ*. Because the GII is a nucleotide barcode sequence that uniquely encodes the

identity of a given gene, the multiplexing capacity increases exponentially as 4^N, where N is the number of sequencing cycles. For example, a GII of length 5 can be read out with 5 sequencing cycles and can detect 4^5 =1,024 distinct genes (although in practice a few extra cycles are used for error correction). This combinatorial coding by sequencing readout thereby allows simultaneous detection of a large number of genes using only a few cycles of imaging (Fig. 1D). Although sequencing readout offers many advantages, BARseq2 is also compatible with hybridization-based readout when necessary. The combination of non-gap-filling *in situ* sequencing of endogenous genes and the gap-filling approach for sequencing barcodes allows many genes to be detected simultaneously with projections using BARseq2.

Our goal is to combine high-throughput projection mapping with single neuron gene profiling to identify cadherin correlates of projections. In the following, we first demonstrate that, by optimizing targeted *in situ* sequencing, BARseq2 can achieve sufficient sensitivity for detection of endogenous mRNAs. We then combine *in situ* sequencing of endogenous mRNAs with *in situ* sequencing of RNA barcodes to associate the expression of cadherins with projection patterns at cellular resolution. We recapitulate previous findings of projection patterns specific to transcriptomic neuronal subtypes and identify cadherins that distinguish major projection classes. We furthermore identify a set of cadherins shared between the mouse auditory cortex and motor cortex that correlate with homologous projections within intratelencephalic (IT) neurons in both cortical areas.

BARseq2 robustly detects endogenous mRNAs

To adequately detect genes using BARseq2, we sought to improve the detection sensitivity. In most *in situ* hybridization methods, high sensitivity is achieved by using many probes for each target mRNA (Chen et al., 2015; Codeluppi et al., 2018; Eng et al., 2019; Raj et al., 2008; Shah et al., 2016; Wang et al., 2018). We reasoned that increasing the number of padlock probes per gene might similarly improve the sensitivity of BARseq2. Indeed, we observed that tiling the whole gene with additional probes (see

Methods for probe design) resulted in as much as a 46-fold increase in sensitivity compared to using a single probe (Fig. 1E; see Supplementary Note 1). Combined with other technical optimizations (Extended Data Fig. 1A, B; see Supplementary Note 1), we increased the sensitivity of BARseq2 to 60 % of RNAscope, a sensitive and commercially available FISH method (Fig. 1F; Extended Data Fig. 1C, D; see Supplementary Note 1). Hence, our optimizations allowed BARseq2 to achieve sufficiently sensitive detection of mRNAs.

To multiplex gene detection with high imaging throughput, we optimized *in situ* sequencing to robustly read out GIIs of single rolonies over many sequencing cycles. We had previously adapted Illumina sequencing chemistry to sequence neuronal somata filled abundantly with RNA barcode rolonies, i.e. DNA nanoballs generated by rolling circle amplification (Chen et al., 2019; Chen et al., 2018). However, directly applying this method to sequence single rolonies produced from individual mRNAs proved difficult due to heating cycles and harsh stripping treatments that led to loss and/or jittering of rolonies (Extended Data Fig. 1E). To allow robust sequencing of single rolonies, we optimized cryo-sectioning (see Methods) and amino-allyl dUTP concentration (Lee et al., 2014) to crosslink rolonies more extensively, achieving less spatial jitter of single rolonies between imaging cycles (Extended Data Fig. 1E-H) and stronger signals (Extended Data Fig. 1I) retained over cycles. This robust *in situ* sequencing of combinatorial GII codes allowed BARseq2 to achieve fast imaging critical for high throughput correlation of gene expression with projections.

BARseq2 allows multiplexed detection of endogenous mRNAs *in situ*

To assess multiplexed detection of cadherins *in situ* using BARseq2, we examined the expression of 20 classical cadherins and non-clustered protocadherins expressed in the adult cortex, along with either three (in auditory cortex) or 45 (in motor cortex) cell-type markers (Fig. 2A-C). In these experiments we used up to 12 probes per gene, which resulted in sensitivity that was sufficient albeit somewhat below the

maximum achievable with more probes. All but three genes were visualized using combinatorial GII codes (4-nt in auditory cortex and 7-nt in motor cortex; see Supp. Table S1); only a small subset of all possible GIIs were used, ensuring a Hamming distance of at least two bases between all pairs of GIIs in auditory cortex and three bases in motor cortex for error correction. The three remaining genes with high expression (*Slc17a7*, *Gad1*, and *Slc30a3*) were detected by hybridization. We successfully resolved and decoded 419,724 rolonies from two slices of mouse auditory cortex (1.7 $mm^2 \times 10$ µm per slice) and 1,445,648 rolonies from four slices of primary motor cortex (2.8 $mm^2 \times 10$ µm per slice). We recovered 20 rolonies in auditory cortex and 115 rolonies in motor cortex that matched two GIIs that were not used in the experiment, corresponding to an estimated error rate of 0.1 % and 0.2 %, respectively, for rolony decoding.

Consistent with previous reports (Basu et al., 2015; Hayano et al., 2014; Krishna et al., 2009; Krishna et al., 2011; Lein et al., 2007; Matsunaga et al., 2015; Paul et al., 2017; Redies, 1997; Tasic et al., 2018), many cadherins were enriched in specific layers and sublayers in the cortex (Fig. 2D). Interestingly, although most cadherins had similar laminar expression in both auditory cortex and motor cortex, some cadherins were differentially expressed across the two areas. For example, *Cdh9* and *Cdh13* were enriched in L2/3 in auditory cortex, but not in motor cortex (Fig. 2D; Extended Data Fig. 2). The laminar positions of peak cadherin expression were consistent with those obtained by other methods, including RNAscope (Fig. 2E; see Supplementary Note 2) and the Allen ISH atlas (Lein et al., 2007)(Fig. 2F; Extended Data Fig. 3, Spearman correlation $\rho = 0.696$ comparing the gene expression density in 100 µm bins across laminae in BARseq2 and in Allen Brain Atlas; see Supplementary Note 2). Thus, BARseq2 accurately resolved the laminar expression patterns of cadherins.

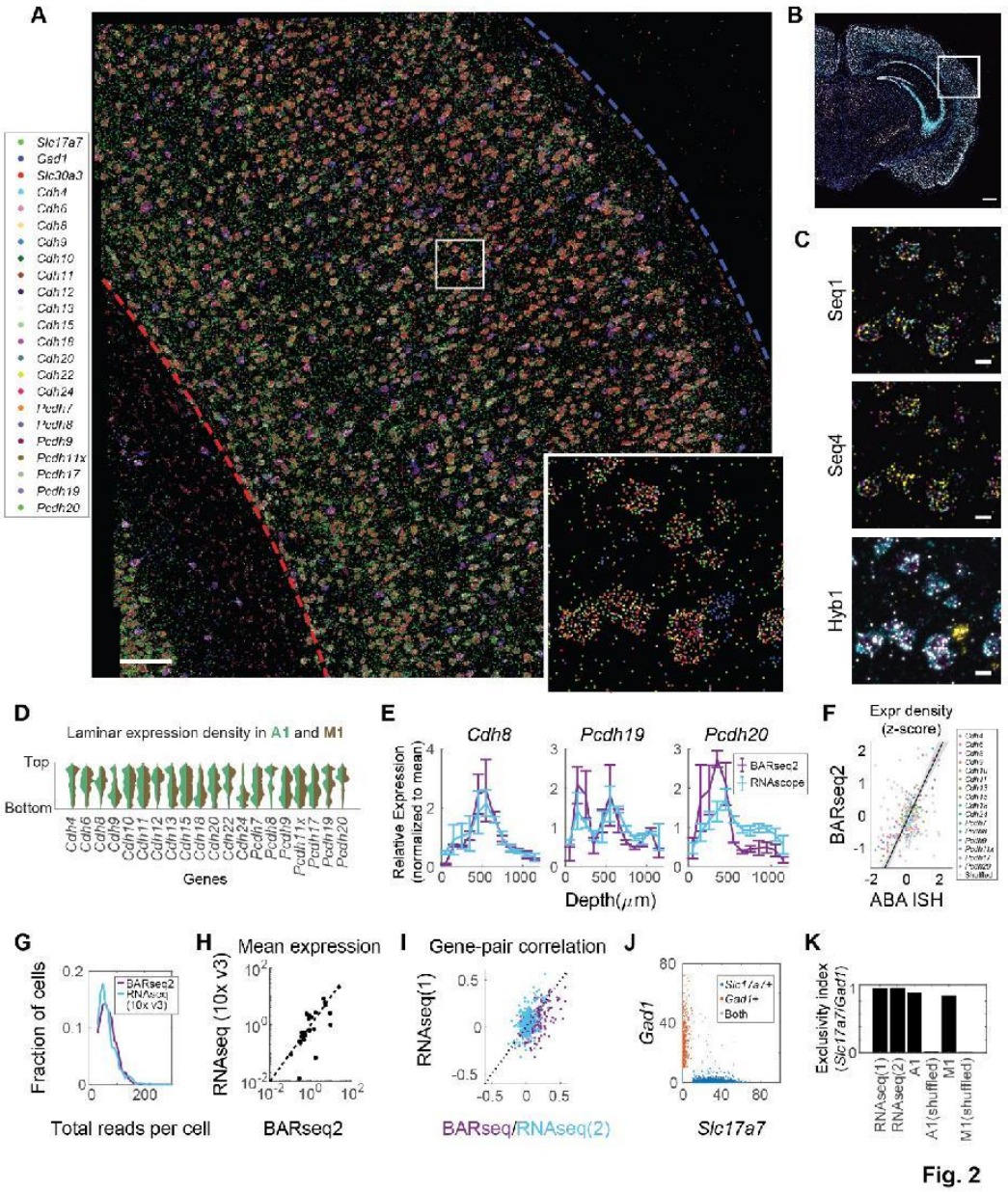

Fig. 2

Figure 2. Multiplexed detection of mRNAs using BARseq2. (A) A representative image of rolonies in auditory cortex. mRNA identities are indicated on the left. The top and bottom of the cortex are indicated by the blue and red dashed lines, respectively. Scale bar = 100 µm. The inset shows a magnified view of the boxed area. (B) Low magnification image of the hybridization cycle showing the location of the area imaged in A. Scale bar = 100 µm (C) Representative images of the indicated sequencing cycle and hybridization cycle of the boxed area in A. Scale bars = 10 µm. (D) Violin plots showing the laminar distribution of cadherin expression in neuronal somata. Expression in auditory cortex and motor cortex is

shown in different colors as indicated. (E) Laminar distribution of *Cdh8*, *Pcdh19*, and *Pcdh20* expression as detected by BARseq2 or FISH, normalized to the mean expression for each gene across all layers. Error bars indicate standard errors. n = 2 slices for BARseq2 and n = 3 slices for FISH. (F) Relative gene expression observed using BARseq2 (y-axis) and in Allen gene expression atlas (x-axis). Each dot represents the expression of a gene in a 100 μm bin in laminar depth. Gene identities are color-coded as indicated. Gray dots indicate correlation between data randomized across laminar positions. A linear fit and 95 % confidence intervals are shown by the diagonal line and the shaded area. (G) Distribution of total read counts per cell in BARseq2 and single-cell RNAseq using 10x v3 in auditory cortex. Only genes used in the panel detected by BARseq2 were included. (H) Mean expression for each gene detected using BARseq2 (x-axis) or single-cell RNAseq (y-axis). Each dot represents a gene. Both axes are plotted in log scale. The dotted line indicates equal expression between BARseq2 and single-cell RNAseq. (I) The correlation between pairs of genes observed in BARseq2 (x-axis, purple dots) and single-cell RNAseq (y-axis, purple dots), or in two single-cell RNAseq datasets (blue dots). The dotted line indicates that correlations between the two datasets are equal. (J) Expression of *Slc17a7* (x-axis) and *Gad1* (y-axis) in single neurons. Color codes indicate whether the neuron dominantly expressed *Slc17a7* (blue) or *Gad1* (red), or expressed both strongly (gray). (K) Exclusivity index (see Methods) of *Slc17a7* and *Gad1* in neurons in two single-cell RNAseq datasets, BARseq2 in auditory or motor cortex, and shuffled BARseq2 data.

We then characterized gene expression obtained by BARseq2 at single-cell resolution. We first defined excitatory and inhibitory neurons as cells having at least 10 reads of either the excitatory marker *Slc17a7* or the inhibitory marker *Gad1,* respectively. We then assigned 228,371 rolonies to 3,377 excitatory or inhibitory neurons [67.6 ± 28.8 (mean ± standard deviation) rolonies per neuron] in auditory cortex, and 752,687 rolonies to 11,492 excitatory or inhibitory neurons [65.5 ± 26.0 (mean ± standard deviation) rolonies per neuron) in motor cortex. Most cadherins showed slight differences in single-cell expression levels in these two cortical areas (Extended Data Fig. 4). In auditory cortex, the total read counts per cell was higher in BARseq2 than in single-cell RNAseq using 10x Genomics v3 (Fig. 2G; median read counts 64 for BARseq2, n = 3,337 cells compared to 57 for single-cell RNAseq, n = 640 cells, $p = 5.3\times10^{-5}$ using rank sum test). Thus, even using a limited number of probes, BARseq2 achieved sensitivity at least equal to single-cell RNAseq using 10x v3. For experiments requiring better quantification of low-expressing genes, the sensitivity could potentially be further improved by using more probes.

Further analyses showed that detection of mRNA by BARseq2 was specific. The mean expression of genes determined by BARseq2 was highly correlated with that determined by single-cell RNAseq using

10x v3 (Fig. 2H; Pearson correlation r = 0.88). Furthermore, correlations between pairs of genes in single neurons determined by BARseq2 were consistent with single-cell RNAseq using 10x v3 to a similar extent as two independent 10x v3 experiments (Fig. 2I; Pearson correlation r = 0.61 and r = 0.52 between BARseq2 and two single-cell RNAseq datasets, respectively, and r = 0.54 between the two single-cell RNAseq datasets; p = 0.78 comparing the difference in correlation between the second single-cell RNAseq dataset to either the first single-cell RNAseq or BARseq2 dataset through bootstrapping). For example, *Slc17a7* and *Gad1*, two genes expressed in either excitatory or inhibitory neurons, respectively, maintained their mutual exclusivity in both auditory cortex and motor cortex as observed by BARseq2 (Fig. 2J, K; see Supplementary Note 3). Similarly, consistent with a previous single-cell RNAseq study (Tasic et al., 2018), BARseq2 also confirmed the observation that *Slc30a3* was more highly expressed in subtypes of excitatory neurons that did not express *Cdh24* compared to projection neurons that did express *Cdh24* [Extended Data Fig. 5A, B; $p = 5 \times 10^{-26}$ using rank sum test on single-cell RNAseq data using Smart-Seq2 (n = 10,044 neurons) (Tasic et al., 2018), and $p = 4 \times 10^{-65}$ on BARseq2 data (n = 2,947 neurons)]. These results indicate that the single-cell gene expression patterns observed by BARseq2 were comparable to those of single-cell RNAseq.

Although finding cadherins that correlate with projections required only low- to medium-level of multiplexing, we wondered if BARseq2 could detect more genes in parallel, and thus be potentially useful in associating projections with larger gene panels. Because imaging time scales logarithmically with the number of genes detected (Fig. 1D), the multiplexing capacity of BARseq2 is limited not by imaging time but by potential reduction in sensitivity when more genes are probed simultaneously. To examine if multiplexing affects detection sensitivity, we probed for *Slc17a7*, *Slc30a3*, and *Gad1* either as a separate three-gene panel or as part of the 65-gene panel (20 cadherins and 45 marker genes). The mean expression density across laminar positions for the three genes were similar between the three-gene panel and the 65-gene panel (Extended Data Fig. 5C; p = 0.22 for *Slc17a7*, p = 0.49 for *Slc30a3*, and p = 0.66 for *Gad1* using rank sum tests, respectively), suggesting that targeting more genes did not affect detection

sensitivity of each gene. Furthermore, the detection of the 65-gene panel in motor cortex (Fig. 3A) allowed us to classify neurons to one of nine transcriptomic neuronal types defined by single-cell RNAseq (Yao et al., 2020) (Fig. 3B; See Supplementary Note 4 and Extended Data Fig. 5D-H). Consistent with previous studies (Tasic et al., 2018; Yao et al., 2020; Zhang et al., 2020), these transcriptomic neuronal types displayed distinct laminar distributions (Fig. 3B, C; See Supplementary Note 4) and cadherin expression (Fig. 3D). These results demonstrate that BARseq2 can be applied to probe gene panels consisting of high dozens and potentially hundreds of genes, with minimal decrease in sensitivity and minimal increase in imaging time.

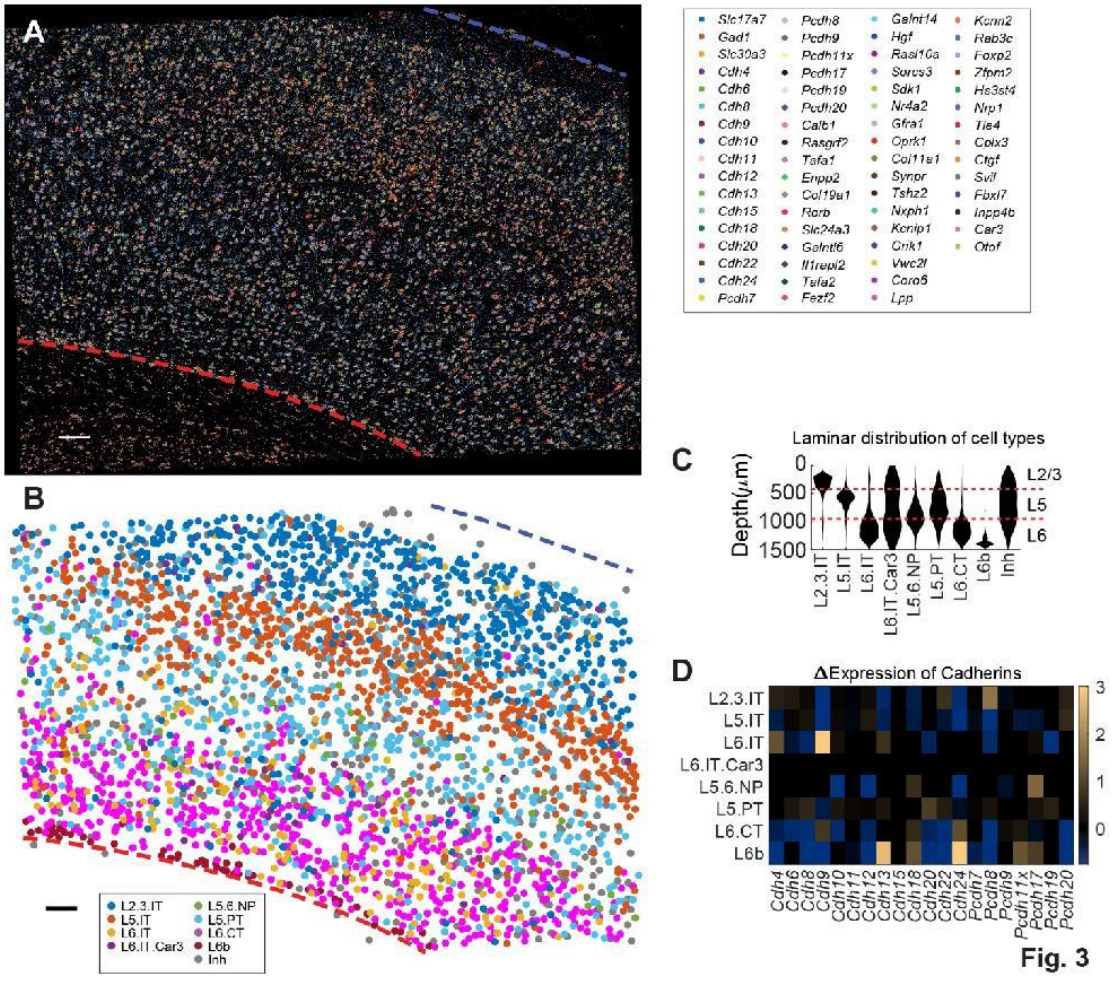

Fig. 3

Figure 3. Cadherin expression across transcriptomic neuronal types in motor cortex. (A) A representative image of rolonies in motor cortex. mRNA identities are color-coded as indicated. The top and the bottom of the cortex are indicated by the blue and red dashed lines, respectively. Scale bar = 100 µm. (B) Transcriptomic cell types called based on gene expression shown in (A). (C) Laminar distribution of transcriptomic neuronal types based on marker gene expression observed by BARseq2. Layer identities are shown on the right. (D) Differential expression of cadherins across transcriptomic neuronal types identified by BARseq2. Over-expression is indicated in yellow and under-expression is indicated in blue. Only differential expression that was statistically significant was shown. Statistical significance was determined using rank sum test with Bonferroni correction for each gene between the indicated transcriptomic type and the expression of that gene across all other neuronal types.

BARseq2 correlates gene expression to projections at cellular resolution with high throughput

To assess cadherin expression and long-range projections in the same cells, we optimized for simultaneous detection and amplification of both endogenous mRNAs and barcodes. Although both endogenous mRNAs and barcodes are amplified using padlock probe-based approaches, amplifying barcodes required the addition of a DNA polymerase to copy barcode sequences into padlock probes to allow direct sequencing of diverse barcodes (up to ~10^{18} diversity; Fig. 1C, *left*). Directly combining the two processes reduced the detection sensitivity of target mRNAs due to the addition of the DNA polymerase [Extended Data Fig. 6A; 37 ± 3 % (mean ± standard error) comparing the *Ctrl* condition to the no polymerase condition]. To preserve detection sensitivity for endogenous mRNAs while allowing the sequencing of diverse barcodes, we adjusted the concentration of the DNA polymerase to 0.001 U/µl (1/200 of the amount in the original BARseq), which doubled the sensitivity for endogenous mRNAs while also maintaining the sensitivity for barcodes (Extended Data Fig. 6A). This optimization allowed BARseq2 to detect both endogenous mRNAs and RNA barcodes together in the same neurons without compromising sensitivity.

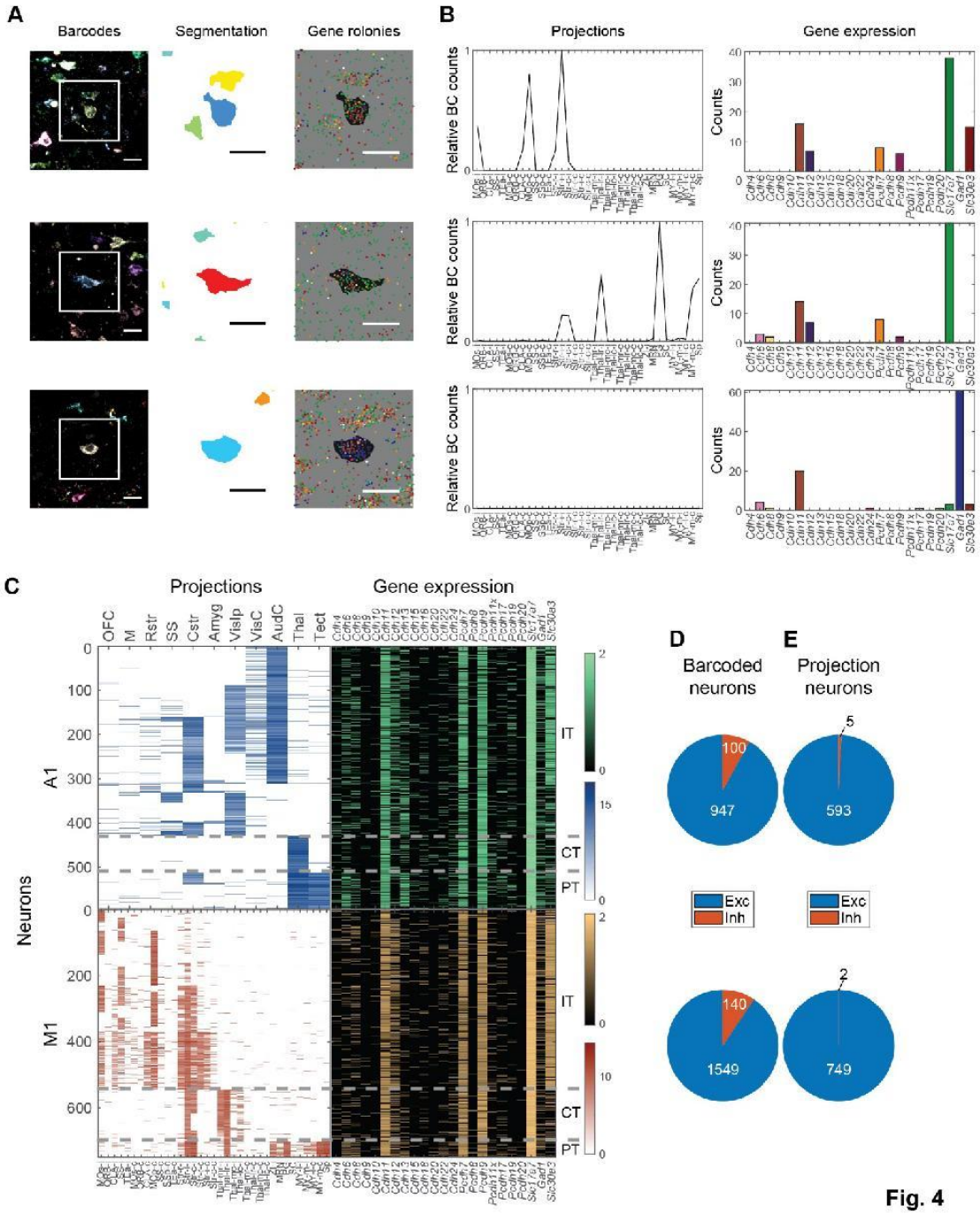

Fig. 4

Figure 4. Correlating gene expression to projections using BARseq2. (A) False-colored barcode sequencing images (*left*), soma segmentations (*middle*), and gene rolonies (*right*) of three representative neurons from the motor cortex. The segmentation and gene rolony images correspond to the white

squared area in the barcode images. In the gene rolony images, the areas corresponding to the soma segmentations of the target neurons are in black. All scale bars = 20 µm. (B) Projections (*left*) and gene expression (*right*) of the target neurons shown in (A). The bars indicating gene expression are colored using the same color code as that in the gene rolony plots in (A). The neurons shown in the first two rows are excitatory projection neurons, whereas the neuron shown in the bottom row is an inhibitory neuron without projections. See Supp. Table S2 for the brain areas corresponding to each abbreviated target area. (C) Projections (*left*) and gene expression (*right*) of neurons in auditory cortex (*top*) and motor cortex (*bottom*). Each row represents a barcoded projection neuron. Both projections and gene expression are shown in log scale. Major projection neuron classes determined by projection patterns are indicated on the right. (D) (E) The number of excitatory neurons (blue) or inhibitory neurons (red) in all barcoded neurons (D) or barcoded projection neurons (E). Neurons in auditory cortex are shown in the top row and those in motor cortex are shown in the bottom row.

We applied BARseq2 to study the correlation between long-range axonal projections and the expression of 20 cadherins, along with three marker genes, in motor cortex and auditory cortex in three mice. In each barcoded cell, we segmented barcoded cell bodies (Fig. 4A, *middle*) using the barcode sequencing images (Fig. 4A, *left*) and assigned rolonies amplified from endogenous genes to the segmented cells (Fig. 4A, *right*). This allowed us to map both the projection patterns of neurons (Fig. 4B, *left*) and gene expression (Fig. 4B, *right*) in the same neurons. To maintain consistency with previous studies (Chen et al., 2019)(Chen et al., unpublished observations), we sampled 11 and 35 projection targets for neurons in auditory cortex and motor cortex, respectively; these projection targets corresponded to most of the major projection targets based on bulk tracing (Oh et al., 2014). We matched barcodes in these target sites to 3,164 well-segmented barcoded neurons (1,283 from auditory cortex and 1,881 from motor cortex) from 15 slices of auditory cortex and 16 slices of motor cortex, each with 10 µm thickness. Of the barcoded neurons, 624 and 791 neurons had projections above the noise floor in auditory cortex and motor cortex, respectively (Fig. 4C). Most neurons [53 % (329/624) in auditory and 89 % (703/791) in motor cortex] projected to multiple brain areas. These observations were largely consistent with previous BARseq experiments in auditory and motor cortex performed without assessing gene expression (Chen et al., 2019)(Chen et al., unpublished observations), confirming that the modifications for BARseq2 did not compromise projection mapping.

BARseq2 recapitulates known projection biases across transcriptomic cell types

Although we have demonstrated that BARseq2 can read out cadherin expression and projections in the same neurons, one might be concerned that barcoding neurons using Sindbis virus could disrupt gene expression (Fros and Pijlman, 2016). To find the relationship between genes and projections, one would require that the gene-gene relationship in Sindbis-infected single neurons reflects that in non-infected neurons, but changes in absolute gene expression level would have little effect. Reassuringly, previous reports have shown that the relationship among genes in single neurons is indeed largely preserved despite a reduction in the absolute expression of genes in Sindbis-infected cells (Chen et al., 2019; Klingler et al., 2018). Furthermore, correlations between transcriptomic types and projections revealed in Sindbis-infected neurons were corroborated by other methods that did not require Sindbis infection (Chen et al., 2019; Wang et al., 2019). Consistent with these previous reports, we observed that the correlations between pairs of genes in the barcoded neurons were consistent with those in non-barcoded neurons (Extended Data Fig. 6B). For example, the expression of the excitatory marker *Slc17a7* and the inhibitory marker *Gad1* remained mutually exclusive in barcoded neurons in both auditory cortex and motor cortex (Extended Data Fig. 6C, D). This mutual exclusivity was preserved despite an overall reduction in mRNA expression (Extended Data Fig. 6E; median reads of 38 in barcoded cells in both auditory and motor cortex, compared to 64 and 48 in non-barcoded cells in the two cortical areas, respectively). Similarly, *Slc30a3* remained differentially expressed across barcoded excitatory neurons with or without *Cdh24* expression as it was in non-barcoded excitatory neurons (Extended Data Fig. 6F; $p = 1 \times 10^{-6}$ using rank sum test, n = 810 neurons). Although our observations cannot rule out the possibility that a small subset of genes may be disrupted by Sindbis infection (e.g. viral response genes), these results suggest that the co-expression relationships of most genes in Sindbis-infected neurons reflect those in non-infected cells. Therefore, the relationship between gene expression and projections resolved by BARseq2 likely reflects that in non-barcoded neurons.

To further test whether BARseq2 can capture the relationship between gene expression and projections, we asked if we could identify differences in projection patterns across transcriptomic neuronal types that could also be validated by previous studies and/or other experimental techniques. We performed these analyses at three different levels of granularity: between excitatory neurons and inhibitory neurons, among major classes of excitatory neurons, and among transcriptomic subtypes of intratelencephalic (IT) neurons in auditory cortex.

First, BARseq2 confirmed that most barcoded neurons with long-range projections were excitatory, not inhibitory. To distinguish between excitatory and inhibitory neurons, we categorized a neuron as excitatory or inhibitory if (1) the neuron had higher expression of the excitatory marker *Slc17a7* or the inhibitory marker *Gad1*, respectively; and (2) the marker was expressed at greater than five reads in the cell. This threshold resulted in 2,496 excitatory neurons (947 in auditory and 1,549 in motor cortex) and 240 inhibitory neurons (100 in auditory cortex and 140 in motor cortex) (Fig. 4D). Consistent with the fact that the majority of long-range projection neurons in the cortex are excitatory, 1,342 of 2,501 (54 %) excitatory neurons, including 593 in auditory cortex and 749 in motor cortex, had detectable projections, whereas only 7 of 240 (3 %) inhibitory neurons (5 in auditory cortex and 2 in motor cortex) had detectable projections (Fig. 4E; see Supplementary Note 5 and Extended Data Fig. 6G, H). The excitatory neurons that did not have detectable projections were likely neurons that projected only locally or to unsampled nearby cortical areas (see Supplementary Note 5). Hence, BARseq2 accurately observed the fact that projection neurons in the cortex are predominantly excitatory and express the excitatory marker *Slc17a7*, not the inhibitory marker *Gad1*.

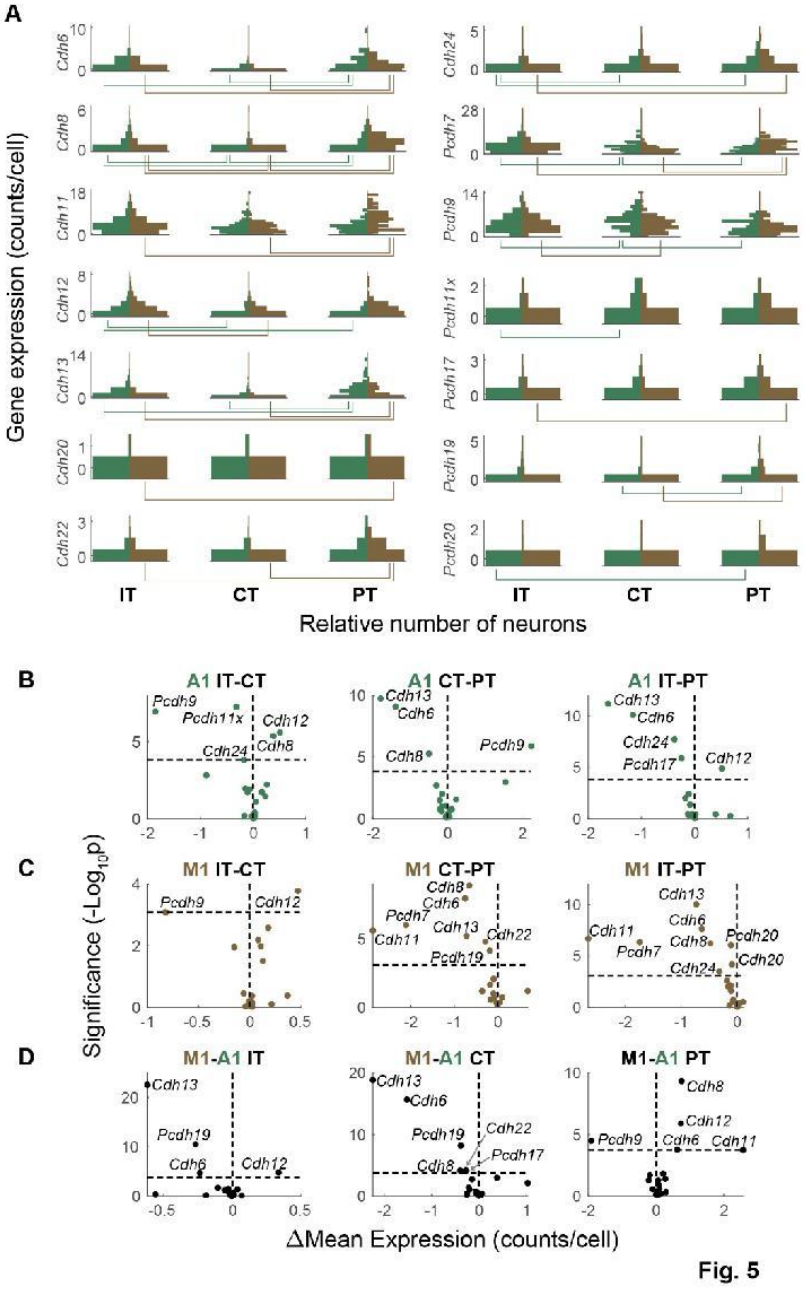

Fig. 5

Figure 5. Differential cadherin expression across major classes and cortical areas. (A) Vertical histograms of the expression (raw counts per cell) of cadherins that were differentially expressed across major classes in either auditory or motor cortex. Y-axes indicate gene expression level (counts per cell) and x-axes indicate number of neurons at that expression level. The numbers of neurons are normalized across plots so that the bins with the maximum number of neurons have equal bar lengths. Gene expression in auditory cortex (green) are shown on the left in each plot, and gene expression in motor

cortex (brown) are shown on the right in each plot. Lines beneath each plot indicate pairs of major classes with different expression of the gene (FDR < 0.05). (B)(C) Volcano plots of cadherins that were differentially expressed across pairs of major classes in auditory cortex (B) or motor cortex (C). Y-axes indicate significance and x-axes indicate effect size. The horizontal dashed lines indicate significance level for FDR < 0.05, and the vertical dashed lines indicate equal expression. (D) Volcano plots of cadherins that were differentially expressed across auditory and motor cortex in the indicated major classes. Y-axes indicate significance and x-axes indicate effect size. The horizontal dashed lines indicate significance level for FDR < 0.05, and the vertical dashed lines indicate equal expression.

Second, BARseq2 revealed differential gene expression across major classes of neurons defined by projections. We found that many cadherins (8 for auditory cortex and 12 for motor cortex) were differentially expressed across intratelencephalic (IT) neurons, pyramidal tract (PT) neurons, and corticothalamic (CT) neurons that were defined by projections (Harris and Shepherd, 2015) (Fig. 5A-C; see Methods for the classification of neurons to projection classes). Several cadherins were consistently differentially expressed in both cortical areas. For example, *Cdh6* and *Cdh13* were over-expressed in PT neurons compared to the other two classes, whereas *Cdh8* was under-expressed in CT neurons compared to the other two classes (FDR < 0.05 using rank sum test). In addition, we also found nine cadherins that were differentially expressed across the two cortical areas in at least one class (Fig. 5D; FDR < 0.05 using rank sum tests). Since IT, PT, and CT neurons can also be classified based on transcriptomic data alone, we compared our findings to the expression of these cadherins observed by single-cell RNAseq (Tasic et al., 2018). The differences in cadherin expression across pairs of classes identified by BARseq were consistent with those observed by single-cell RNAseq (Extended Data Fig. 7A; the rank correlation of the differences in cadherin expression across major neuronal types was 0.61 between BARseq and single-cell RNAseq, compared to 0.39 between auditory and motor cortex in BARseq; see Supplementary Note 6). Thus, BARseq2 identified cadherin correlates of major neuronal classes that were consistent with those observed using single-cell RNAseq.

Finally, BARseq2 confirmed known biases in projection patterns across transcriptomic IT subtypes in auditory cortex (Extended Data Fig. 7B, C). Previous studies using both barcoding-based strategy and

single-cell tracing have identified distinctive projection patterns for two transcriptomic subtypes of IT neurons, IT3 (L6 IT) and IT4 (L6 *Car3*+ IT) (Chen et al., 2019; Wang et al., 2019). To test if we could capture the same projection specificity of transcriptomic subtypes, we mapped projection patterns to projection clusters identified in a previous study in auditory cortex, and used a combination of gene expression and laminar position to distinguish four transcriptomic subtypes of IT neurons (Chen et al., 2019)(see Methods for detailed definitions). As expected, the two transcriptomic subtypes (IT3 and IT4) predominantly found in L5 and L6 were indeed more likely to project only to the ipsilateral cortex, without projections to the contralateral cortex or the striatum ($p = 4 \times 10^{-7}$ comparing the fraction of neurons with only ipsilateral cortical projections in IT3/IT4 to the fraction of them in IT1/IT2 using Fisher's test; Extended Data Fig. 7B, C). Between IT3 and IT4, IT4 neurons were more likely to project ipsilaterally (58 % IT3 neurons compared to 92 % IT4 neurons, $p = 1 \times 10^{-4}$ using Fisher's test), whereas IT3 neurons were more likely to project contralaterally (66 % IT3 neurons compared to 14 % IT4 neurons, $p = 5 \times 10^{-8}$ using Fisher's test). Thus, BARseq2 recapitulated known projection differences across transcriptomic subtypes of IT neurons.

BARseq2 identifies cadherin correlates of IT projections

Having established that BARseq2 identified gene correlates of projections that were consistent with previous studies, we then moved to identify cadherins that correlate with projections within IT neurons. We first grouped projections to different areas based on the likelihood of co-innervation, and then identified genes that correlated with these groups of projections. The projections of IT neurons to multiple brain areas correlated with each other in both auditory cortex and motor cortex (Chen et al., 2019)(Fig. 6A). For example, neurons in the auditory cortex that projected to the somatosensory cortex (SS) were also more likely to project to the ipsilateral visual cortex (VisIp), but not the contralateral auditory cortex (AudC). To exploit these correlations, we used non-negative matrix factorization (NMF)(Lee and Seung,

1999), an algorithm related to PCA but with the added constraint that projections be non-negative, to represent the projection pattern of each neuron as the sum of several "projection modules." Each of these modules (six modules for the motor cortex and three for the auditory cortex; Fig. 6B) consisted of subsets of projections that were likely to co-occur. We named these modules by the main projections (cortex, CTX, or striatum, STR) followed by the side of the projection (ipsilateral, -I, or contralateral, -C). For some modules, we further indicated that the projections were to the caudal part of the structure by prefixing with "C" (e.g. CSTR-I or CCTX-I). A small number of projection modules could explain most of the variance in projections (three modules and six modules explained 84 % and 87 % of the variance in projections to nine areas in auditory cortex and 18 areas in motor cortex that IT neurons project to, respectively; Fig. 6C).

We found many cadherins whose expression co-varied with projection modules (Supp. Fig. S1). For example, auditory cortex neurons expressing *Pcdh19* were stronger in the CSTR-I projection module than those not expressing *Pcdh19* [Fig. 6D, *top*; $p = 5 \times 10^{-4}$ comparing the CSTR-I module in neurons with (n = 83) or without (n = 346) *Pcdh19* expression using rank sum test]. Surprisingly, the same association between *Pcdh19* and the CSTR-I projection module was also seen in motor cortex (Fig. 6D, *bottom*; $p = 4 \times 10^{-6}$ using rank sum test, n = 31 for *Pcdh19*+ neurons and n = 512 for *Pcdh19*- neurons), despite the overall differences in projections from these two areas. Similarly, *Cdh8* was correlated with the CTX-I module and *Cdh12* was correlated with the CTX-C module (Fig. 6E, FDR < 0.1) in both auditory and motor cortex. These correlations were independently validated by retrograde tracing using cholera toxin subunit B (CTB) and FISH (Extended Data Fig. 8A-E; See Supplementary Note 7). Additional cadherins, including *Cdh6*, *Cdh11*, *Cdh20*, *Pcdh7*, and *Pcdh9*, were also correlated with projection modules in at least one of the two areas (Fig. 6E, FDR < 0.1; Supp. Fig. S1). Our observations that the same cadherins correlated with similar projection modules in both areas suggest that a common molecular logic might underscore the organization of projections across cortical areas beyond class-level divisions.

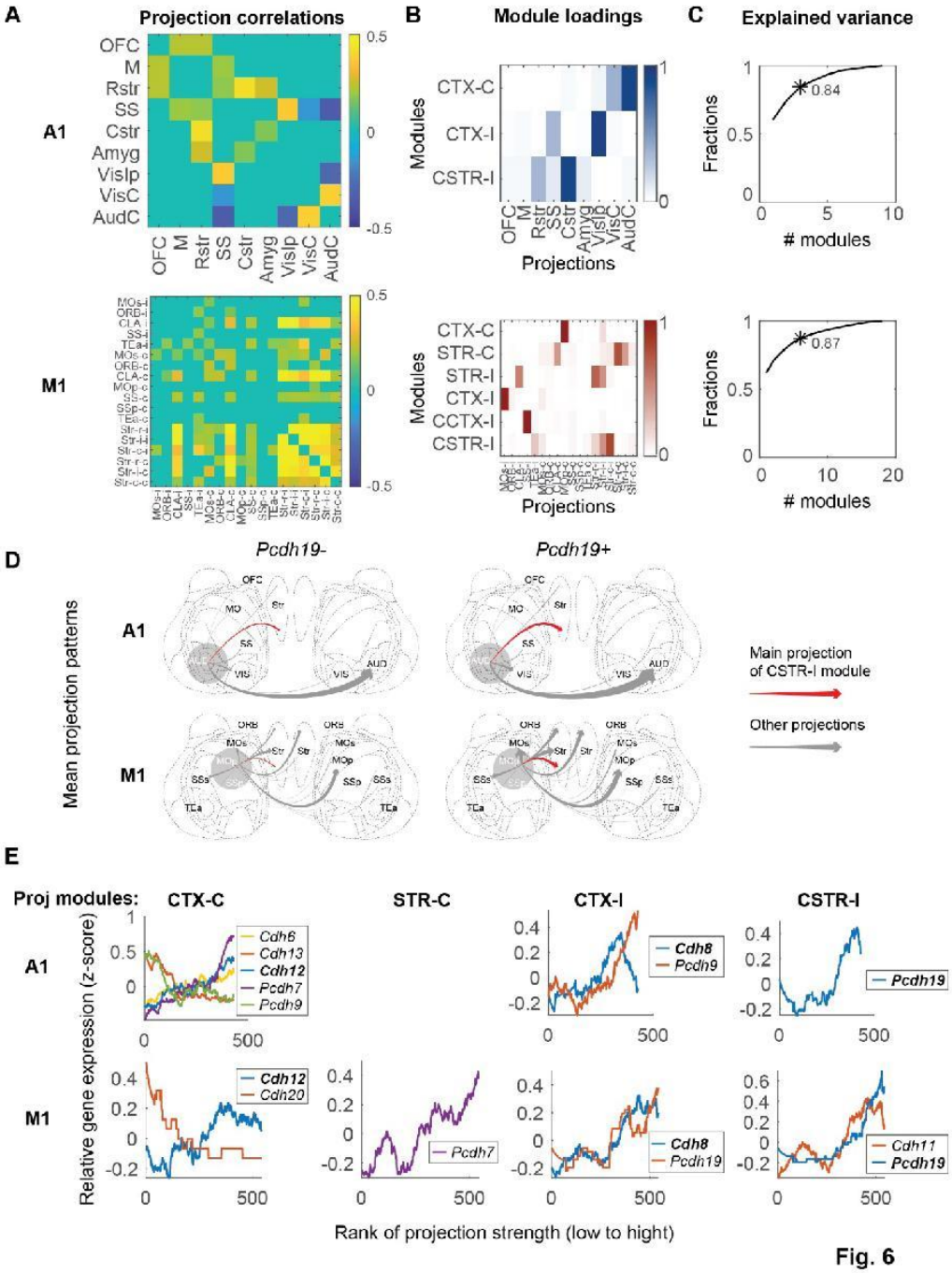

Figure 6. Cadherins correlate with diverse projections of IT neurons. (A) Correlation of projections to different brain areas in IT neurons of auditory cortex (*top*) or motor cortex (*bottom*). Only significant correlations are shown. (B) Projection modules of IT neurons in auditory cortex (*top*) or motor cortex

(*bottom*). Each row represents a projection module. Columns indicate projections to different brain areas. (C) The fractions of variance explained by different numbers of projection modules in auditory cortex (*top*) and motor cortex (*bottom*). The numbers of projection modules that correspond to those in (B) are labeled with an asterisk with the fraction of variance explained indicated. (D) Mean projection patterns of neurons in A1 (*top*) and M1 (*bottom*) with or without *Pcdh19* expression. The thickness of arrows indicates projection strength (barcode counts). Red arrows indicate projections that correspond to the strongest projection in the CSTR-I projection modules. (E) The expression of cadherins (y-axes) that were rank correlated with the indicated projection modules in auditory cortex (*top row*) and motor cortex (*bottom row*). Neurons (x-axes) are sorted by the strengths of the indicated projection modules. Only genes that were significantly correlated with projection modules are shown (FDR < 0.1). Genes that were correlated with the same projection modules in both areas are shown in bold.

Although gene expression is inherently noisy, the expression of many genes is correlated in single neurons. We reasoned that the correlations among genes might allow us to identify additional relationships between gene expression and projections that were missed by analyzing each gene separately. To exploit the correlations among genes, we grouped 16 cadherins into three meta-analytic co-expression modules based on seven single-cell RNAseq datasets of IT neurons in motor cortex (Fig. 7A; Extended Data Fig. 9A, B) (Yao et al., 2020). To obtain the modules, we followed the rank-based network aggregation procedure defined by Ballouz et al. (2015) and Crow et al. (2016) to combine the seven dataset-specific gene-gene co-expression networks into an aggregated network, and then grouped together genes showing consistent excess correlation using the dynamic cutting tree algorithm (Langfelder et al., 2008). Two co-expressed modules were associated with projections: Module 1 was associated with contralateral striatal projections (STR-C projection module), and Module 2 was associated with ipsilateral caudal striatal projections (CSTR-I; Fig. 7B, C; Extended Data Fig. 9C, D). These associations between the co-expression modules and projections were consistent with, but stronger than, associations between individual genes contained in each module and the same projections (Extended Data Fig. 9E). Interestingly, these co-expression modules were enriched in multiple transcriptomic subtypes of IT neurons, but these transcriptomic subtypes were found in multiple branches of the transcriptomic taxonomy (Fig. 7D; Extended Data Fig. 9F). For example, Module 1 is associated with transcriptomic subtypes of IT neurons in L2/3, L5, and L6. This result is consistent with previous

observations (Chen et al., 2019; Tasic et al., 2018; Zhang et al., 2020) that first-tier transcriptomic subtypes of IT neurons (i.e. subtypes of the highest level in the transcriptomic taxonomy within the IT class) appeared to share projection patterns, and further raises the possibility that transcriptomic taxonomy does not necessarily capture differences in projections. Taken together, our finding that projections correlate with cadherin co-expression modules independent of transcriptomic subtypes demonstrates that BARseq2 can reveal intricate relationships between gene expression and projection patterns.

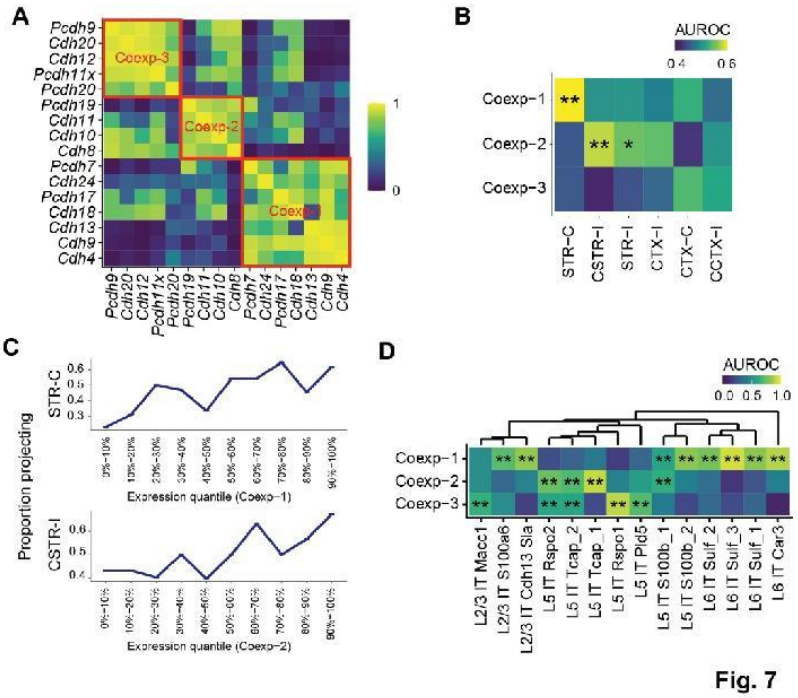

Fig. 7

Figure 7. Gene co-expression modules correlate with diverse projections of IT neurons. (A) Correlation among cadherins as identified using single-cell RNAseq in IT neurons in motor cortex (Yao et al., 2020). Three co-expression modules are marked by red squares. Cadherins that did not belong to any module were not shown. (B) Association between cadherin co-expression modules and projection modules (AUROC). Significant associations are marked by asterisks (*FDR < 0.1, **FDR < 0.05). (C) Fractions of neurons with the indicated projection modules as a function of co-expression module expression. Neurons are binned by gene module quantiles as indicated. (D) Association of the three co-expression modules in transcriptomic IT neurons in the scSS dataset (AUROC, significance shown as in B).

Discussion

BARseq2 combines high-throughput mapping of projections to many brain areas with multiplexed detection of gene expression at single-cell resolution. Because BARseq2 is high-throughput, we are able to correlate gene expression and projection patterns of thousands of individual neurons in a single experiment, and thereby achieve statistical power that would be challenging to obtain using other single-cell techniques. By applying BARseq2 to two distant cortical areas—primary motor and auditory cortex—in the adult mouse, we identified cadherin correlates of diverse projections. Our results suggest that BARseq2 provides a path to discovering general organization of gene expression and projections that are shared across the cortex.

BARseq2 detects multiplexed gene expression with high throughput

To correlate panels of genes, such as cadherins, to projections, we designed BARseq2 to detect gene expression with high throughput, multiplex to dozens of genes, have sufficient sensitivity, and be compatible with barcoding-based projection mapping. To satisfy these needs, we based BARseq2 on padlock probe-based approaches (Ke et al., 2013; Qian et al., 2020). With additional optimizations for sensitivity, sequencing readout, and compatibility with barcode sequencing, we successfully used BARseq2 to identify cadherins that correlate with projections.

One of the critical requirements for BARseq2 is high throughput when reading out many genes. Through strong amplification of mRNAs, combinatorial coding, and robust readout using Illumina sequencing chemistry (Chen et al., 2019; Chen et al., 2018), BARseq2 achieves fast imaging at low optical resolution compared to many other imaging-based spatial transcriptomic methods (Chen et al., 2015; Codeluppi et al., 2018; Eng et al., 2019; Shah et al., 2016; Zhang et al., 2020). Further

optimizations, including computational approaches for resolving spatially mixed rolonies (Chen et al., 2020), have the potential to increase imaging throughput even further.

Another critical optimization was increasing the low sensitivity that early versions of the padlock probe-based technique suffered from, unless special and expensive primers were used (Ke et al., 2013). Inspired by other spatial transcriptomic methods, we and others (Qian et al., 2020) have found that tiling target genes with multiple probes could greatly improve the sensitivity. This design allowed variable sensitivity for different experimental purposes. Although we identified cadherin correlates of projections using a modest number of probes per gene to achieve sensitivity similar to single-cell RNAseq using 10x Genomics v3, we could achieve much higher sensitivity using more probes. This high and tunable sensitivity, combined with the fact that the gene multiplexing capacity of BARseq2 is not limited by imaging time, opens potential applications of BARseq2 to a wide range of questions that require high-throughput interrogation of gene expression *in situ*.

BARseq2 reveals gene correlates of projections

BARseq2 exploits the high-throughput axonal projection mapping that BARseq offers to identify gene correlates of diverse projections. BARseq has sensitivity comparable to single neuron tracing (Han et al., 2018). Although the spatial resolution of BARseq for projections is lower than conventional single neuron tracing, it offers throughput that is several orders of magnitude higher than the state-of-the-art single-cell tracing techniques (Lin et al., 2018; Winnubst et al., 2019). This high throughput allows BARseq to reveal higher-order statistical structure in projection patterns that would have been difficult to observe using existing techniques, such as single-cell tracing (Chen et al., 2019; Kebschull et al., 2016). The increased statistical power of BARseq, obtained at the cost of some spatial resolution, is reminiscent of different clustering power across single-cell RNAseq techniques of varying throughput and read depth

(Ding et al., 2020; Yao et al., 2020). The high throughput of BARseq thereby provides a powerful asset for investigating the organization of projection patterns and their relationship to gene expression.

BARseq2 enables simultaneous measurement of multiplexed gene expression and axonal projections to many brain areas, at single neuron resolution and at a scale that would be difficult to achieve with other approaches. For example, *Cre*-dependent labeling allows interrogation of the gene expression and projection patterns of a genetically defined subpopulation of neurons (Chen et al., 2019). However, this approach lacks cellular resolution and is limited by the availability of *Cre* lines and requires that a neuronal population of interest be specifically distinguished by the expression of one or two genes. The combination of single-cell transcriptomic techniques with retrograde labeling does provide cellular resolution, but can only interrogate projections to one or at most a small number of brain areas at a time (Economo et al., 2018; Kim et al., 2019; Tasic et al., 2018; Zhang et al., 2020). The inability to interrogate projections to many brain areas from the same neuron would miss higher-order statistical structures in projections, which are non-random (Han et al., 2018) and provide additional information regarding other properties of the neurons, such as laminar position and gene expression (Chen et al., 2019)(Chen et al., unpublished observations). The projections of individual neurons to multiple brain areas can be obtained using multiplexed single-cell tracing (Lin et al., 2018; Wang et al., 2019; Winnubst et al., 2019), but the throughput of these methods remains relatively low. Moreover, many advanced single-cell tracing techniques require special sample processing that hinders multiplexed interrogation of gene expression in the same sample. BARseq2 thus provides a powerful tool for probing the relationships between gene expression and projection patterns.

Cadherins correlate with diverse projections of IT neurons.

Using BARseq2, we identified several cadherins that correlate with homologous IT projections in both auditory and motor cortex, two spatially and transcriptomically distant areas with distinct cortical and subcortical projection targets. We speculate that these cadherin correlates of projections shared across areas may represent the remnants of a common developmental program that establishes similar projections (Custo Greig et al., 2013), or may be needed for ongoing functions or maintenance of projections. Our findings raise the possibility that a shared molecular code might underlie the diversity of cortical projections.

Although the cadherin correlates of projections were observed in adult neurons and thus did not necessarily reflect the genes that specified projections during development, our observations are compatible with the "cadherin code hypothesis" (Duan et al., 2014; Duan et al., 2018; Krishna et al., 2011; Redies, 1997), according to which combinations of cadherins can specify complex projection patterns. Although recent studies have provided evidence of cadherins in specifying diverse connectivity in the retina (Duan et al., 2014; Duan et al., 2018), this hypothesis remains difficult to test systematically in more complex and heterogeneous circuits such as the cortex due to a lack of high throughput techniques with high multiplexing capacity for gene detection. BARseq2 possesses the high throughput and multiplexing qualities essential for investigating such questions, and could potentially be applied to circuit development to illuminate genetic and anatomical changes over developmental trajectories. BARseq2 thus provides a path to discovering the myriad genetic programs that specify and/or correlate with long-range projections in both developing and mature animals.

BARseq2 builds a unified description of neuronal diversity

Recent developments in methodologies have enabled the high-throughput interrogation of individual neuronal properties, such as connectivity, neuronal activity, genomic signatures, and developmental

lineage, at cellular resolution. However, a comprehensive understanding of neuronal circuits further requires the examination of how distinct neuronal characteristics relate to each other. It is essential to correlate different neuronal properties at cellular resolution to further our understanding of neuronal circuitry and function.

BARseq2 represents an important step toward this goal by integrating multiplexed gene expression and neuronal projections to many brain areas at cellular resolution with high throughput. Although we focused on the relationship between cadherin expression and projections, our results suggest a complex relationship between transcriptomic cell types and projections: the cadherin co-expression modules that correlated with projections were enriched in multiple transcriptomic clusters across different branches of the transcriptomic taxonomy. This observation raises the interesting possibility that the hierarchy of transcriptomic cell types does not necessarily capture distinctions in long-range connectivity, a hypothesis that is consistent with several previous studies (Chen et al., 2019; Tasic et al., 2018; Zhang et al., 2020).

Because BARseq2 integrates neuronal properties using spatial information, it is potentially compatible with other *in situ* assays, such as immunohistochemistry, two-photon calcium imaging, and dendritic morphological reconstruction. BARseq2 can additionally characterize genetic labeling tools, such as *Cre*-lines (Chen et al., 2019), to potentially allow genetic access and manipulation of neuronal subpopulations. Interrogating and manipulating neuronal populations in the context of various neuronal properties with high throughput and cellular resolution can elucidate the fundamental rules that govern organization of cortical neuronal diversity. By spatially correlating various neuronal properties in single neurons, BARseq2 represents a feasible path towards achieving a comprehensive description of neuronal circuits.